

ABSTRAK

Masalah kualitas udara telah menjadi permasalahan global karena dampaknya yang signifikan terhadap kesehatan manusia dan lingkungan. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi kualitas udara dengan menggunakan metode machine learning. Proses pemilihan fitur dilakukan untuk mengidentifikasi parameter yang paling berpengaruh, seperti karbon monoksida (CO), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), suhu, serta kedekatan dengan area industri dan pengukuran lainnya. Dalam penelitian ini, efisiensi model diukur menggunakan metrik seperti Area Under the Curve (AUC) dan Matthews Correlation Coefficient (MCC). Dua algoritma utama yang digunakan adalah K-Nearest Neighbors (KNN) dan Decision Tree. Hasil penelitian menunjukkan bahwa model Decision Tree memberikan performa terbaik dengan nilai AUC sebesar 0,9380 dan akurasi klasifikasi (CA) sebesar 0,8820. Sementara itu, KNN juga menunjukkan peningkatan signifikan setelah penerapan metode seleksi fitur. Penerapan metode ini tidak hanya meningkatkan efisiensi analisis data, tetapi juga menghasilkan prediksi yang lebih akurat. Penelitian ini membuka peluang untuk pengembangan lebih lanjut, termasuk penerapan metode lain atau kombinasi tambahan parameter guna mengatasi tantangan yang lebih kompleks di masa depan.

Kata Kunci: Prediksi Kualitas Udara, Seleksi Fitur, Information Gain, K-Nearest Neighbors, Decision Tree